

A Nonlinear Entropic Variational Model for Image Filtering

A. Ben Hamza

Concordia Institute for Information Systems Engineering, Concordia University, Montréal, Quebec H3G 1T7, Canada
Email: hamza@ciise.concordia.ca

Hamid Krim

Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, NC 27695-7911, USA
Email: ahk@ncsu.edu

Josiane Zerubia

Ariana Research Group, INRIA/I3S, BP 93, 06902 Sophia Antipolis Cedex, France
Email: josiane.zerubia@inria.fr

Received 12 August 2003; Revised 8 June 2004

We propose an information-theoretic variational filter for image denoising. It is a result of minimizing a functional subject to some noise constraints, and takes a hybrid form of a negentropy variational integral for small gradient magnitudes and a total variational integral for large gradient magnitudes. The core idea behind this approach is to use geometric insight in helping to construct regularizing functionals and avoiding a subjective choice of a prior in maximum a posteriori estimation. Illustrative experimental results demonstrate a much improved performance of the approach in the presence of Gaussian and heavy-tailed noise.

Keywords and phrases: MAP estimation, variational methods, robust statistics, differential entropy, gradient descent flows, image denoising.

1. INTRODUCTION

In recent years, variational methods and partial differential equations-(PDE) based methods [1, 2, 3, 4, 5, 6] have been introduced to explicitly account for intrinsic geometry to address a variety of problems including image segmentation, mathematical morphology, motion estimation, image classification, and image denoising [7, 8, 9, 10, 11, 12]. The latter will be the focus of the present paper. The problem of signal/image denoising has been addressed using a number of different techniques including wavelets [13], order statistics-based filters [14], PDE-based algorithms [9, 15], and variational approaches [16, 17, 18]. In particular, a large number of PDE-based methods have particularly been proposed to tackle the problem of image denoising [12, 19, 20] with a good preservation of edges. Much of the appeal of PDE-based methods lies in the availability of a vast arsenal of mathematical tools which at the very least act as a key guide in achieving numerical accuracy as well as stability. PDEs or gradient descent flows are generally a result of variational problems using the Euler-Lagrange principle [21]. One popular

variational technique used in image denoising is the total variation-based approach. It was developed in [4] to overcome the basic limitations of all smooth regularization algorithms, and a variety of numerical methods have also recently been developed for solving total variation minimization problems [22, 23].

In this paper, we present a variational approach to maximum a posteriori (MAP) estimation. The core idea behind this approach is to use geometric insight in helping construct regularizing functionals and avoiding a subjective choice of a prior in MAP estimation. Using tools from robust statistics and information theory, we show that we can extend this strategy and develop two gradient descent flows for image denoising with a demonstrated performance.

In Section 3, a general formulation of signal/image denoising problem is stated. In Section 3, we briefly recall the MAP estimation technique, and in Section 4 we formulate the problem of MAP estimation in the calculus of variations setting. Section 5 is devoted to a robust variational formulation using concepts borrowed from robust estimation, followed by a probabilistic interpretation of the nonlinear

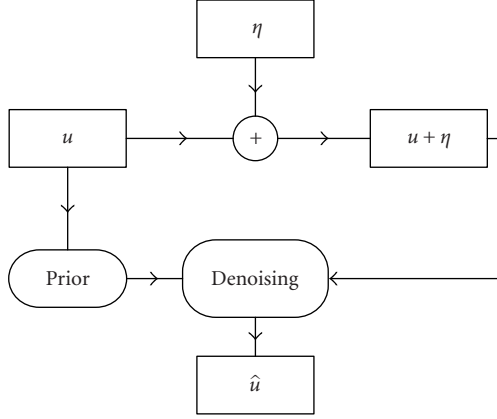


FIGURE 1: Block diagram of image denoising process.

anisotropic diffusion in the order statistics framework. In Section 6, information-theoretic variational flows based on the differential entropy are proposed. In Section 7, we provide experimental results to demonstrate a much improved performance of the proposed gradient descent flows in image denoising. Finally, some conclusions and discussions are included in Section 8.

2. PROBLEM STATEMENT

In all real applications, measurements are perturbed by noise. In the course of acquiring, transmitting, or processing a digital image for example, the noise-induced degradation may be dependent on or independent of data. The noise is usually described by its probabilistic model, for example, Gaussian noise is characterized by two moments. Application-dependent, a degradation often yields a resulting signal/image observation model, and the most commonly used is the additive one

$$u_0 = u + \eta, \quad (1)$$

where the observed image u_0 includes the original signal u and the independent and identically distributed (i.i.d) noise process η .

Image denoising refers to the process of recovering an image contaminated by noise (see Figure 1). The challenge of the problem of interest lies in faithfully recovering the underlying signal/image u from u_0 , and furthering the estimation by making use of any prior knowledge/assumptions about the noise process η . This goal is graphically and succinctly described in Figure 1.

3. MAP ESTIMATION: MODEL-BASED APPROACH

In a probabilistic setting, the image denoising problem is usually solved in a discrete domain, and in this case, an image is expressed by a random matrix $u = (u_{ij})$ of gray levels. To account for prior probabilistic information we may have for u , a technique of choice is that of a MAP estimation. Denoting by $p(u)$ the prior distribution for the unknown image u ,

the MAP estimator is given by

$$\hat{u} = \arg \max_u \{ \log p(u_0|u) + \log p(u) \}, \quad (2)$$

where $p(u_0|u)$ denotes the conditional probability of u_0 given u .

A general model for the prior distribution $p(u)$ is that of a Markov random field (MRF) which is characterized by its Gibbs distribution given by [24]

$$p(u) = \frac{1}{Z} \exp \left\{ - \frac{\mathcal{F}(u)}{\lambda} \right\}, \quad (3)$$

where Z is a partition function and λ is a constant known as the temperature in the terminology of physical systems. \mathcal{F} is called the *energy function* and has the form $\mathcal{F}(u) = \sum_{c \in \mathcal{C}} V_c(u)$, where \mathcal{C} denotes a set of cliques (i.e., a set of connected pixels) for the MRF, and V_c is a potential function defined on a clique. We may define the cliques to be adjacent pairs of horizontal and vertical pixels. Note that for large λ , the prior probability becomes flat, and for small λ , the prior probability exhibits sharp modes.

MRFs have been extensively used in computer vision particularly for image restoration, and it has been established that the Gibbs distributions and MRFs are equivalent (see e.g., [24]). In other words, if a problem is defined in terms of local potentials, then there is a simple way of formulating the problem in terms of MRFs. If the noise process η is i.i.d. Gaussian, then we have

$$p(u_0|u) = K \exp \left(- \frac{|u - u_0|^2}{2\sigma^2} \right), \quad (4)$$

where K is a normalizing positive constant, σ^2 is the noise variance, and $|\cdot|$ stands for the Euclidean norm or for the absolute value in the case of a scalar. Thus, the MAP estimator in (2) yields

$$\hat{u} = \arg \min_u \left\{ \mathcal{F}(u) + \frac{\lambda}{2} |u - u_0|^2 \right\}. \quad (5)$$

Image estimation using MRF priors has proven to be a powerful approach to restoration and reconstruction of high-quality images. Its major drawback, besides its computational load, is the difficulty in systematically selecting a practical and reliable prior distribution. The Gibbs prior parameter λ is also of particular importance since it controls the balance of influence of the Gibbs prior and that of the likelihood. If λ is too small, the prior will tend to have an over-smoothing effect on the solution. Conversely, if it is too large, the MAP estimator may be unstable and it reduces to the maximum likelihood solution as λ goes to infinity. Another difficulty in using a MAP estimator is the nonuniqueness of the solution when the energy function \mathcal{F} is not convex.

4. A VARIATIONAL APPROACH TO MAP ESTIMATION

Unknown prevailing statistics or underlying signal/image/noise models often make a “target” desired performance

quantitatively less well defined. Specifically, it may be qualitative in nature (e.g., preserve high gradients in a geometric setting, or determine a worst-case noise distribution in a statistical estimation setting with a number of interpretations), and may not necessarily be tractably assessed by an objective and optimal performance measure. The formulation of such qualitative goals is typically carried out by way of adapted functionals which upon being optimized achieve the stated goal, for example, a monotonically decreasing functional of gradient modifying a diffusion [2]. This approach is the so-called *variational* approach. It is commonly formulated in a continuous domain which enjoys a large arsenal of analytical tools, and hence offers a greater flexibility. An image is therefore defined as a real-valued function $u : \Omega \rightarrow \mathbb{R}$, and Ω is a nonempty, bounded, open set in \mathbb{R}^2 (usually Ω is a rectangle in \mathbb{R}^2). Throughout, $\mathbf{x} = (x_1, x_2)$ denotes a pixel location in Ω , and $\|\cdot\|$ denotes the L^2 -norm. While the ultimate overall objective in the aforementioned formulation may coincide with that of a probabilistic formulation, namely, the recovery of an underlying desired signal u , it is herein often implicit and embedded in an energy functional to be optimized. Generally, the construction of an energy functional is based on some characteristic quantity specified by the task at hand (gradient for segmentation, Laplacian for smoothing, etc.). This energy functional is oftentimes coupled to a regularizing force/energy in order to rule out a great number of solutions and to also avoid any degenerate solution.

When considering the signal model (1), our goal may be succinctly stated as one of estimating the underlying image u based on an observation u_0 and/or any potential knowledge of the noise statistics to further regularize the solution. This yields the fidelity-constrained optimization problem

$$\begin{aligned} \min_u \mathcal{F}(u) \\ \text{s.t. } \|u - u_0\|^2 = \sigma^2, \end{aligned} \quad (6)$$

where \mathcal{F} is a given functional which often defines, as noted above, the particular emphasis on the features of the achievable solution. In other words, we want to find an optimal solution that yields the smallest value of the objective functional among all solutions that satisfy the constraints. Using Lagrange's theorem, the minimizer of (6) is given by

$$\hat{u} = \arg \min_u \left\{ \mathcal{F}(u) + \frac{\lambda}{2} \|u - u_0\|^2 \right\}, \quad (7)$$

where λ is a nonnegative parameter chosen so that the constraint $\|u_0 - u\|^2 = \sigma^2$ is satisfied. In practice, the parameter λ is often estimated or chosen a priori.

Equations (5) and (7) show a close connection between image recovery via MAP estimation and image recovery via optimization of variational integrals. One may in fact reexpress (5) in an integral form similar to that of (7).

A critical issue, however, is the choice of the variational integral \mathcal{F} , which as discussed later, is often driven by geometric arguments. Among the better known functionals (also called *variational integrals*) in image denoising are the Dirichlet and the total variation integrals defined, respectively, as

$$\mathcal{D}(u) = \frac{1}{2} \int_{\Omega} |\nabla u|^2 d\mathbf{x}, \quad \text{TV}(u) = \int_{\Omega} |\nabla u| d\mathbf{x}, \quad (8)$$

where ∇u denotes the gradient of the image u . The total variation method basically consists in finding an estimate \hat{u} for the original image u with the smallest total variation among all the images satisfying the noise constraint $\|u - u_0\|^2 = \sigma^2$, where σ is assumed known. Note that the parameter λ controls the tradeoff between noise removal and detail preservation.

The intuition for the use of this variational integral is that it incorporates the fact that discontinuities are present in the original image u (it measures the jumps of u even if it is discontinuous). The total variation method has been used with success in image denoising, especially for denoising images with piecewise constant features while preserving the location of the edges exactly [19].

A generalization of the Dirichlet and total variation functionals is the variational integral given by

$$\mathcal{F}(u) = \int_{\Omega} F(|\nabla u|) d\mathbf{x}, \quad (9)$$

where $F : \mathbb{R}^+ \rightarrow \mathbb{R}$ is a given smooth function called a *variational integrand* or *Lagrangian* [21]. Using (9), we hence define a functional

$$\begin{aligned} \mathcal{L}(u) &= \mathcal{F}(u) + \frac{\lambda}{2} \|u - u_0\|^2 \\ &= \int_{\Omega} \left(F(|\nabla u|) + \frac{\lambda}{2} \|u - u_0\|^2 \right) d\mathbf{x}, \end{aligned} \quad (10)$$

which by the formulation in (7) becomes

$$\hat{u} = \arg \min_{u \in X} \mathcal{L}(u), \quad (11)$$

where X is an appropriate image space of smooth functions like $C^1(\overline{\Omega})$, or the space $BV(\Omega)$ of image functions with bounded variation, or the Sobolev space $H^1(\Omega) = W^{1,2}(\Omega)$. Note that $BV(\Omega)$ is a Banach space with the norm $\|u\|_{BV} = \|u\|_{L^1(\Omega)} + \text{TV}(u)$, while $H^1(\Omega)$ is a Hilbert space with the norm $\|u\|_{H^1(\Omega)}^2 = \|u\|^2 + \|\nabla u\|^2$.

4.1. Properties of the optimization problem

A problem is said to be *well posed* in the sense of Hadamard if (i) a solution of the problem exists, (ii) the solution is unique, and (iii) the solution is stable, that is, depends continuously on the problem data. It is *ill-posed* when it fails to satisfy at least one of these criteria. To guarantee the well-posedness of our minimization problem (11), the following result provides some conditions.

Proposition 1. *Let the image space X be a reflexive Banach space, and let \mathcal{F} be*

- (i) *weakly lower semicontinuous, that is, if for any sequence (u^k) in X converging weakly to u , $\mathcal{F}(u) \leq \liminf_{k \rightarrow \infty} \mathcal{F}(u^k)$;*
- (ii) *coercive, that is, $\mathcal{F}(u) \rightarrow \infty$ as $\|u\| \rightarrow \infty$.*

Then the functional \mathcal{L} is bounded from below and possesses a minimizer; that is, there exists $\hat{u} \in X$ such that $\mathcal{L}(\hat{u}) = \inf_X \mathcal{L}$. Moreover, if \mathcal{F} is convex and $\lambda > 0$, then the optimization problem (11) has a unique solution, and therefore it is stable.

Proof. From (i) and (ii) and the weak lower semicontinuity of the L^2 -norm, the functional \mathcal{L} is weak lower semicontinuous, and coercive, that is, $\mathcal{L}(u) \rightarrow \infty$ as $\|u\| \rightarrow \infty$.

Let u^n be a minimizing sequence of \mathcal{L} , that is, $\mathcal{L}(u^n) \rightarrow \inf_X \mathcal{L}$. An immediate consequence of the coercivity of \mathcal{L} is that u^n must be bounded. As X is reflexive, thus u^n converges weakly to \hat{u} in X , that is, $u^n \rightharpoonup \hat{u}$. Thus $\mathcal{L}(\hat{u}) \leq \liminf_{n \rightarrow \infty} \mathcal{L}(u^n) = \inf_X \mathcal{L}$. This proves that $\mathcal{L}(\hat{u}) = \inf_X \mathcal{L}$.

It is easy to check that convexity implies weakly lower semicontinuity. Thus the solution of the optimization problem (11) exists and it is unique because the L^2 -norm is strictly convex. The stability follows using the semicontinuity of \mathcal{L} and the fact that u^n is bounded. \square

4.2. Numerical solution: gradient descent flows

To solve the optimization problem (11), a variety of iterative methods such as gradient descent [4] or fixed point method [22, 23] may be applied.

The first-order necessary condition to be satisfied by any minimizer of the functional \mathcal{L} given by (10) is that its first variation $\delta\mathcal{L}(u; v)$ vanishes at u in direction of v , that is,

$$\delta\mathcal{L}(u; v) = \left. \frac{d}{d\epsilon} \mathcal{L}(u + \epsilon v) \right|_{\epsilon=0} = 0, \quad (12)$$

and a solution u of (12) is called a *weak extremal* of \mathcal{L} [21].

Using (10) and (12), we obtain the first variation $\delta\mathcal{L}(u; v)$ (see Appendix A for a proof):

$$\begin{aligned} \delta\mathcal{L}(u; v) &= \int_{\Omega} \left\{ \left(\frac{F'(|\nabla u|)}{|\nabla u|} \nabla u \cdot \nabla v \right) + \lambda(u - u_0)v \right\} d\mathbf{x} \\ &= - \int_{\Omega} \left\{ \operatorname{div} \left(\frac{F'(|\nabla u|)}{|\nabla u|} \nabla u \right) + \lambda(u - u_0) \right\} v d\mathbf{x} \end{aligned} \quad (13)$$

for all $v \in X$.

Using the fundamental lemma of the calculus of variations, this vanishing first variation yields an Euler-Lagrange equation as a necessary condition to be satisfied by minimizers of \mathcal{L} . In mathematical terms, the Euler-Lagrange equation is given by

$$- \operatorname{div} \left(\frac{F'(|\nabla u|)}{|\nabla u|} \nabla u \right) + \lambda(u - u_0) = 0 \quad \text{in } \Omega, \quad (14)$$

where “div” stands for the divergence operator. An image u satisfying (14) is called an *extremal* of \mathcal{L} .

Note that $|\nabla u|$ is not differentiable when $\nabla u = 0$ (e.g., flat regions in the image u). To overcome the resulting numerical difficulties, we use the slight modification

$$|\nabla u|_{\epsilon} = \sqrt{|\nabla u|^2 + \epsilon}, \quad (15)$$

where ϵ is positive sufficiently small.

Proposition 2. *Let $\lambda = 0$ and let S be a convex set of an image space X . If the Lagrangian F is nonnegative convex and of class C^1 , then every weak extremal of \mathcal{L} is a minimizer of \mathcal{L} on S .*

Proof. The convexity of F yields

$$F(y) \geq F(x) + F'(x)(y - x) \quad \forall x, y \in \mathbb{R}^+. \quad (16)$$

By assumption, u is a weak extremal of \mathcal{L} , that is, $\delta\mathcal{L}(u; v) = 0$ for all $v \in S$. This implies that $F'(|\nabla u|) = 0$. Therefore, using (16), we obtain

$$\int_{\Omega} F(|\nabla v|) d\mathbf{x} \geq \int_{\Omega} F(|\nabla u|) d\mathbf{x}. \quad (17)$$

This concludes the proof. \square

By further constraining λ , we may be in a position to sharpen the properties of the minimizer, as given in the following.

Proposition 3. *Let $\lambda = 0$ and let S be a convex set of an image space X . If the Lagrangian F is nonnegative convex and of class C^1 such that $F'(0) \geq 0$, then the global minimizer of \mathcal{L} is a constant image.*

Proof. Using (16), it follows that $F(|\nabla u|) \geq F(0)$. Thus the constant image is a minimizer of \mathcal{L} . Since S is convex, it follows that this minimizer is global. \square

Proposition 4. *Let $\lambda > 0$ and let S be a convex set of image space X . If the Lagrangian F is nonnegative, strictly convex, and of class C^1 , then an extremal u of \mathcal{L} is the unique minimizer of \mathcal{L} on S .*

Proof. Since $u \mapsto (\lambda/2)|u - u_0|^2$ is strictly convex when $\lambda > 0$, then the functional $\mathcal{L}(u)$ is strictly convex on S , that is,

$$\mathcal{L}(v) > \mathcal{L}(u) + \nabla \mathcal{L}(u) \cdot (v - u). \quad (18)$$

By assumption, u is an extremal of \mathcal{L} , thus $\mathcal{L}(v) > \mathcal{L}(u)$ for all $v \neq u$. \square

Using the Euler-Lagrange variational principle, the minimizer of (11) may be interpreted as the steady-state solution to the following nonlinear elliptic PDE called *gradient descent flow*:

$$u_t = \operatorname{div} (g(|\nabla u|) \nabla u) - \lambda(u - u_0) \quad \text{in } \Omega \times \mathbb{R}_+, \quad (19)$$

where $g(z) = F'(z)/z$, with $z > 0$ and assumed homogeneous Neumann boundary conditions. A numerical implementation of this PDE is discussed in Appendix B.



FIGURE 2: Image evolution of (a) the original image under (b) the heat flow, (c) Perona-Malik flow, and (d) the total variation flow.

4.3. Illustrative cases

The following examples illustrate the close connection between optimization problems of variational integrals and boundary value problems for PDEs in a *no-noise* constraint case (i.e., setting $\lambda = 0$).

Heat equation

$u_t = \Delta u$ is the gradient descent flow for the Dirichlet variational integral $\mathcal{D}(u)$.

It is important to point out that the Dirichlet functional tends to smooth out sharp jumps because it controls the second derivative of image intensity, that is, its “spatial acceleration,” and it diffuses the intensity values isotropically. Figure 2b shows this blurring effect on a clean image depicted in Figure 2a.

Perona-Malik (PM) equation

It has been shown in [25] that the PM diffusion $u_t = \text{div}(g(|\nabla u|)\nabla u)$ is the gradient descent flow for the variational integral

$$\mathcal{F}_c(u) = \int_{\Omega} F_c(|\nabla u|) d\mathbf{x}, \quad (20)$$

with sample Lagrangians $F_c^1(z) = c^2 \log(1 + z^2/c^2)$ or $F_c^2(z) = c^2(1 - \exp(-z^2/c^2))$, where $z \in \mathbb{R}^+$ and c is a tuning positive constant. These Lagrangians are depicted in Figure 3.

A minimization of such functionals encourages the smoothing of homogeneous/small gradient regions and the preservation of edges/high-gradient regions. Note that ill-posedness of this formulation was addressed in a number of papers (e.g., see [25]). A result of applying the PM flow

with F_c^1 to the original image in Figure 2a is illustrated in Figure 2c. It is worth noting how the diffusion takes place throughout the homogeneous regions and not across the edges.

Curvature flow

$u_t = \text{div}(\nabla u/|\nabla u|)$ corresponds to the total variation integral.

While limiting spurious oscillations, total variation optimization preserves sharp jumps as is often encountered in “blocky” signals/images. Figure 2d illustrates the output of the total variation flow.

5. ROBUST VARIATIONAL APPROACH

5.1. Robustness for unknown statistics

In robust estimation, for example, a case where even the noise statistics are *not* precisely known [13, 26] arises. In this case, a reasonable strategy would be to assume that the noise is a member of some set, or of some class of parametric families, and to pick the worst-case density (*least favorable* in some sense) member of that set, and obtain the best signal reconstruction for it. Huber’s ϵ -contaminated normal set \mathcal{P}_ϵ is defined as [26]

$$\mathcal{P}_\epsilon = \{(1 - \epsilon)\Phi + \epsilon H : H \in \mathcal{H}\}, \quad (21)$$

where Φ is the standard normal distribution, \mathcal{H} is the set of all probability distributions symmetric with respect to the origin, and $\epsilon \in [0, 1]$ is the known fraction of “contamination.” Huber found that the least favorable distribution in \mathcal{P}_ϵ which maximizes the asymptotic variance (or, equivalently, minimizes the Fisher information) is Gaussian in the center

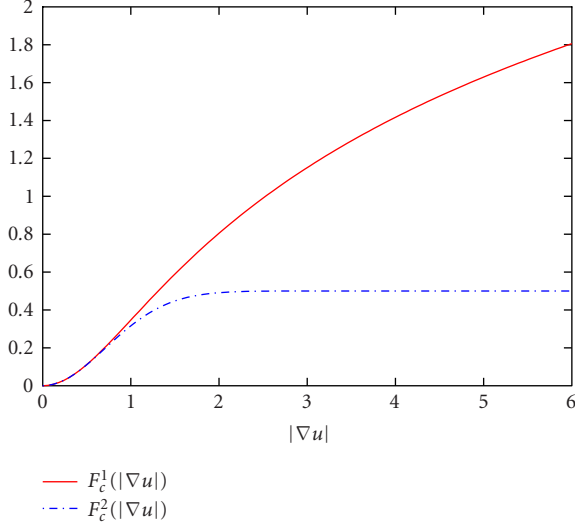


FIGURE 3: Anisotropic Lagrangians.

and Laplacian in the tails. The transition between the two depends on the fraction of contamination ϵ , that is, larger fractions correspond to smaller switching points and vice versa.

For the set \mathcal{P}_ϵ of ϵ -contaminated normal distributions, the least favorable distribution has a density function

$$f_H(z) = \left(\frac{1-\epsilon}{\sqrt{2\pi}} \right) \exp(-\rho_k(z)), \quad (22)$$

where ρ_k is the Huber M -estimator cost function (see Figure 4) given by

$$\rho_k(z) = \begin{cases} \frac{z^2}{2} & \text{if } |z| \leq k, \\ k|z| - \frac{k^2}{2} & \text{otherwise.} \end{cases} \quad (23)$$

Here k is a positive constant determined by the fraction of contamination ϵ by the equation

$$2 \left(\frac{\phi(k)}{k} - \Phi(-k) \right) = \frac{\epsilon}{1-\epsilon}, \quad (24)$$

where Φ is the standard normal distribution function and ϕ is its probability density function. It is clear that ρ_k is a convex function, quadratic in the center and linear in the tails as illustrated in Figure 4.

Motivated by the robustness of the Huber M -filter in a probabilistic setting and its resilience to impulsive noise, we propose a variational filter which, when accounting for these properties, leads to the energy functional

$$\mathcal{R}_k(u) = \int_{\Omega} \rho_k(|\nabla u|) d\mathbf{x}. \quad (25)$$

Note that the Huber variational integral is a hybrid of the Dirichlet variational integral ($\rho_k(|\nabla u|) \propto |\nabla u|^2/2$ as $k \rightarrow \infty$) and of the total variation integral ($\rho_k(|\nabla u|) \propto |\nabla u|$ as $k \rightarrow 0$). One may check that the Huber variational integral

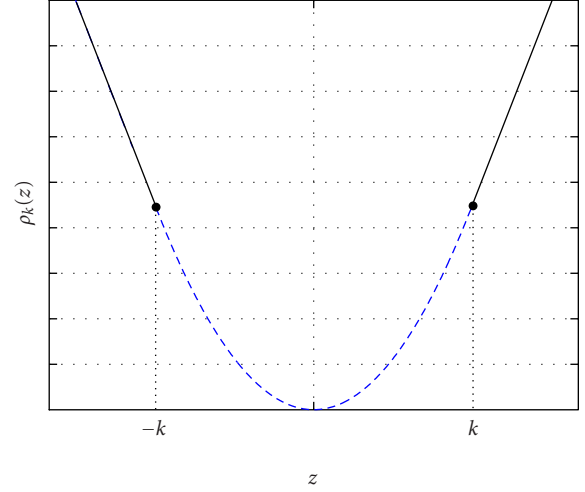


FIGURE 4: Huber function.

$\mathcal{R}_k : H^1(\Omega) \rightarrow \mathbb{R}^+$ is well defined, convex, and coercive. It follows from Proposition 1 that the minimization problem

$$\hat{u} = \arg \min_{u \in H^1(\Omega)} \left\{ \mathcal{R}_k(u) + \frac{\lambda}{2} \|u - u_0\|^2 \right\} \quad (26)$$

$$= \arg \min_{u \in H^1(\Omega)} \int_{\Omega} \left(\rho_k(|\nabla u|) + \frac{\lambda}{2} |u - u_0|^2 \right) d\mathbf{x}$$

has a solution. This solution is unique when $\lambda > 0$.

Proposition 5. *The optimization problem (26) is equivalent to*

$$\hat{u} = \arg \min_{(u, \theta) \in H^1(\Omega) \times \mathbb{R}} \left\{ \frac{\theta^2}{2} + \int_{\Omega} \left(k| |\nabla u| - \theta | + \frac{\lambda}{2} |u - u_0|^2 \right) d\mathbf{x} \right\}. \quad (27)$$

Proof. For z fixed, define $\Psi(\theta) = (1/2)\theta^2 + k|z - \theta|$ on \mathbb{R} . It is clear that Ψ is convex on \mathbb{R} . It follows that Ψ attains its minimum at θ_0 such that $\Psi'(\theta_0) = 0$ and $\Psi''(\theta_0) > 0$, that is, $\theta_0 = k \operatorname{sign}(z - k)$. Thus we have

$$\Psi(\theta_0) = \begin{cases} kz - \frac{k^2}{2} & \text{if } z > k, \\ \frac{z^2}{2} & \text{if } z = k, \\ -kz - \frac{k^2}{2} & \text{if } z < -k, \end{cases} \quad (28)$$

and therefore $\rho_k(z) = \arg \min_{\theta \in \mathbb{R}} \Psi(\theta)$. This concludes the proof. \square

Using the Euler-Lagrange variational principle, a Huber gradient descent flow is obtained as

$$u_t = \operatorname{div} (g_k(|\nabla u|) \nabla u) - \lambda(u - u_0) \quad \text{in } \Omega \times \mathbb{R}_+, \quad (29)$$

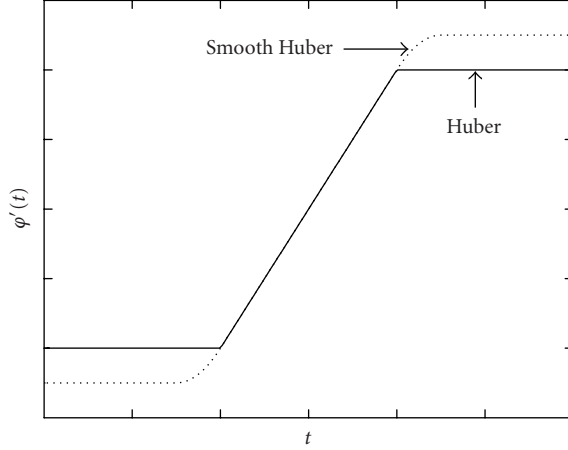


FIGURE 5: Huber influence function and its smooth version.

where g_k is the Huber M -estimator weight function

$$g_k(z) = \frac{\rho'_k(z)}{z} = \begin{cases} 1 & \text{if } |z| \leq k, \\ \frac{k}{|z|} & \text{otherwise.} \end{cases} \quad (30)$$

For large k , this flow yields an isotropic diffusion (heat equation when $\lambda = 0$), and for small k , it corresponds to the total variation gradient descent flow (curvature flow when $\lambda = 0$).

It is worth pointing out that in the case of no-noise constraint (i.e., setting $\lambda = 0$), the Huber gradient descent flow yields a robust anisotropic diffusion [27] obtained by replacing the diffusion functions proposed in [2] with robust M -estimator weight functions [26].

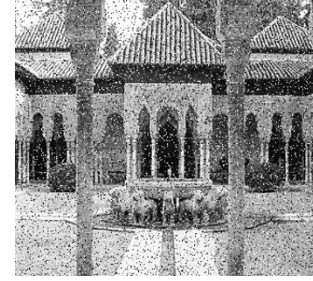
Recently, we proposed a smooth Huber variational integral [28] given by

$$\Phi(u) = \int_{\Omega} \varphi(|\nabla u|) d\mathbf{x}, \quad (31)$$

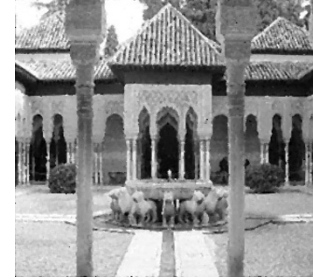
where the Lagrangian φ is defined as

$$\varphi(t) = \begin{cases} -ct & \text{if } t \leq -a, \\ \frac{(t+a)^3}{3} - ct & \text{if } -a < t < -b, \\ \frac{t^2 - b^2}{2} + \frac{(a-b)^3 + 3bc}{3} & \text{if } -b \leq t \leq b, \\ \frac{-(t-a)^3}{3} + ct & \text{if } b < t < a, \\ ct & \text{otherwise,} \end{cases} \quad (32)$$

with $a = 3/2$, $b = 1$, and $c = 5/4$. Its derivative φ' (also referred to as an influence function in robust statistics) is depicted in Figure 5. The Huber influence function, however, is not differentiable as shown in Figure 5. The differentiability of the influence function is of great importance since it implies the continuity of its first derivative which in turn implies the continuity of the confidence intervals in the data points. A more detailed description of the smooth Huber gradient descent flow will be reported elsewhere.



(a)



(b)

FIGURE 6: Log-Cauchy filtering: (a) contaminated image with impulsive noise; (b) filtered image.

5.2. Perona-Malik equation: an estimation-theoretic perspective

In a similar spirit as above, one may proceed to justify the PM equation from a specific statistical model. Assuming an image $u = (u_{ij})$ as a random matrix with i.i.d. elements, the output of the log-Cauchy filter [14] is defined as a solution to the maximum log-likelihood estimation problem for a Cauchy distribution with dispersion c and estimation parameter θ . In other words, the output of a log-Cauchy filter is the solution to the robust estimation problem [14]

$$\min_{\theta} \sum_{i,j} \log \left(c^2 + (u_{ij} - \theta)^2 \right) = \min_{\theta} \sum_{i,j} F_c(u_{ij} - \theta), \quad (33)$$

where the cost function F_c coincides with the Lagrangian function which yields the PM equation. Hence, in the probabilistic setting, the PM flow corresponds to the log-Cauchy filter. Figure 6 illustrates the performance of the log-Cauchy filter in removing heavy-tailed (impulsive) noise.

6. INFORMATION-BASED FUNCTIONALS

6.1. Information-theoretic approach

In the previous section, we proposed a *least favorable* distribution as a result of exercising *our ignorance* in describing that of an image gradient within a domain. Another effective way is to adopt a criterion which bounds such a case,

namely, that of *entropy* [29]. The maximum entropy criterion is indeed an important principle in statistics for modeling the prior probability $p(u)$ of a process u , and has been used with success in numerous image processing applications. The term is often associated with qualifying the selection of a distribution subject to some moments constraints (e.g., mean, variance, etc.), that is, the available information is described by way of moments of some known functions $m_r(u)$ with $r = 1, \dots, s$. Indeed, coupling the finiteness of $m_r(u)$ for example with the maximum entropy condition of the data suggests a *most random* model $p(u)$ with the corresponding moments constraints as a most adapted model (equivalently minimizing; negentropy, see, e.g., [30]):

$$\begin{aligned} \min_u \int p(u) \log p(u) du \\ \text{s.t. } \int p(u) du = 1, \\ \int m_r(u) p(u) du = \mu_r, \quad r = 1, \dots, s. \end{aligned} \quad (34)$$

Using Lagrange's theorem, the solution of (34) is given by

$$p(u) = \frac{1}{Z} \exp \left\{ - \sum_{r=1}^s \lambda_r m_r(u) \right\}, \quad (35)$$

where λ_r 's are the Lagrange multipliers, and Z is a partition function. The resulting model $p(u)$ given by (35) may hence be used as a prior in a MAP estimation formulation.

6.2. Entropic gradient descent flow

Motivated by the good performance of the maximum entropy principle in image/signal analysis applications and inspired by its rationale, we may naturally adapt it to describe the distribution of a gradient throughout an image. Specifically, the large gradients should coincide with tail events of this distribution, while the small and medium ones representing the smooth regions form the mass of the distribution. Towards that end, we write

$$\mathcal{H}(u) = \int_{\Omega} H(|\nabla u|) d\mathbf{x} = \int_{\Omega} |\nabla u| \log |\nabla u| d\mathbf{x}, \quad (36)$$

where $H(z) = z \log(z)$, $z \geq 0$. Note that $-H(z) \rightarrow 0$ as $z \rightarrow 0$.

It follows from the inequality $z \log(z) \leq z^2/2$ that

$$|\mathcal{H}(u)| \leq \int_{\Omega} |\nabla u|^2 d\mathbf{x} \leq \|u\|_{H^1(\Omega)}^2 < \infty \quad \forall u \in H^1(\Omega), \quad (37)$$

where $\|\cdot\|_{H^1(\Omega)}$ denotes the H^1 -norm. Thus the negentropy variational integral $\mathcal{H} : H^1(\Omega) \rightarrow \mathbb{R}$ is well defined. Note also that the inequality $z \log(z) \leq z^2/2$ implies $\mathcal{H}(u) \leq \mathcal{D}(u)$, where $\mathcal{D}(u)$ is the Dirichlet integral. One may check that the Lagrangian H is strictly convex, and coercive, that is, $H(z) \rightarrow \infty$ as $|z| \rightarrow \infty$. The following result follows from Proposition 1.

Proposition 6. *Let $\lambda > 0$. The minimization problem*

$$\begin{aligned} \hat{u} &= \arg \min_{u \in H^1(\Omega)} \left\{ \mathcal{H}(u) + \frac{\lambda}{2} \|u - u_0\|^2 \right\} \\ &= \arg \min_{u \in H^1(\Omega)} \int_{\Omega} \left(|\nabla u| \log |\nabla u| + \frac{\lambda}{2} |u - u_0|^2 \right) d\mathbf{x} \end{aligned} \quad (38)$$

has a unique solution provided that $|\nabla u| \geq 1$.

Calling upon the Euler-Lagrange variational principle again, the entropic gradient descent flow is given by

$$u_t = \operatorname{div} \left(\frac{1 + \log |\nabla u|}{|\nabla u|} \nabla u \right) - \lambda(u - u_0), \quad \text{in } \Omega \times \mathbb{R}_+, \quad (39)$$

with homogeneous Neumann boundary conditions. In addition, this energy spread of the gradient energy may be related to that sought by the total variation method, which in contrast allows for additional higher gradients.

Proposition 7. *Let u be an image. The negentropy variational integral and the total variation satisfy the following inequality:*

$$\mathcal{H}(u) \geq TV(u) - 1. \quad (40)$$

Proof. Since the negentropy H is a convex function, the Jensen inequality yields

$$\begin{aligned} \int_{\Omega} H(|\nabla u|) d\mathbf{x} &\geq H \left(\int_{\Omega} |\nabla u| d\mathbf{x} \right) \\ &= H(TV(u)) \\ &= TV(u) \log TV(u), \end{aligned} \quad (41)$$

and using the inequality $z \log(z) \geq z - 1$ for $z \geq 0$, we conclude the proof. \square

Remark 1 (see Figure 7). The following inequalities between Huber variational, integral, total variation, and negentropy integral hold:

- (i) if $|\nabla u| \in [0, e]$, then $\mathcal{H}(u) \leq TV(u)$;
- (ii) if $|\nabla u| \in (e, \infty)$, then $\mathcal{H}(u) > TV(u)$;
- (iii) if $|\nabla u| \in (e^{k-1}, \infty)$ and $k \geq 2$, then $\mathcal{H}(u) \leq \mathcal{R}_k(u)$,

where e is the Euler number ($e = \lim_{n \rightarrow \infty} (1 + 1/n)^n \approx 2.71$).

6.3. Improved entropic gradient descent flow

To summarize and for a comparison sake, we show in Figure 7 the behavior of the variational integrands we have discussed in this paper. It can be readily shown [18] that a differentiable hybrid functional between the negentropy variational integral and the total variation may be defined as

$$\mathcal{H}_{TV}(u) = \begin{cases} \mathcal{H}(u) & \text{if } |\nabla u| \leq e, \\ 2TV(u) - \operatorname{meas}(\Omega)e & \text{otherwise,} \end{cases} \quad (42)$$

yielding an improved gradient descent flow. The quantity $\operatorname{meas}(\Omega)$ denotes the Lebesgue measure of the image

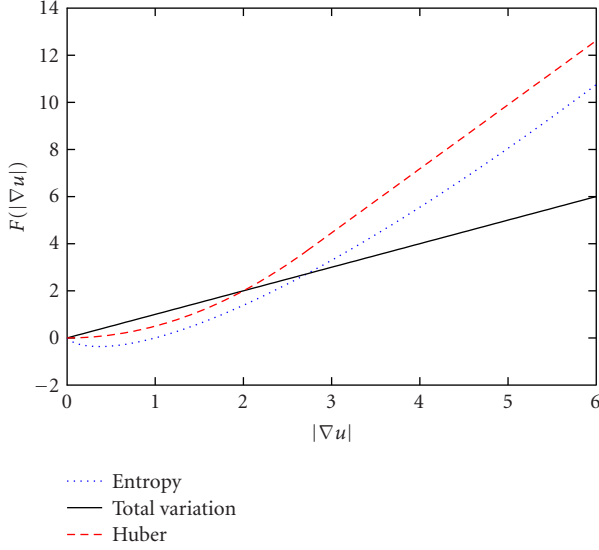


FIGURE 7: Visual comparison of some variational integrands.

domain Ω . Note that $\mathcal{H}_{TV} : H^1(\Omega) \rightarrow \mathbb{R}$ is well defined, differentiable, convex, and coercive. It follows from Proposition 1 that the minimization problem

$$\hat{u} = \arg \min_{u \in H^1(\Omega)} \left\{ \mathcal{H}_{TV}(u) + \frac{\lambda}{2} \|u - u_0\|^2 \right\} \quad (43)$$

has a unique solution provided that $\lambda > 0$.

Figure 8 depicts the improved entropic Lagrangian $H_{TV} : \mathbb{R}^+ \rightarrow \mathbb{R}$ defined as

$$H_{TV}(z) = \begin{cases} z \log(z) & \text{if } z \leq e, \\ 2z - e & \text{otherwise} \end{cases} \quad (44)$$

Using the Euler-Lagrange variational principle, it follows that the improved entropic gradient descent flow is given by

$$u_t = \nabla \cdot \left(\frac{H'_{TV}(|\nabla u|)}{|\nabla u|} \nabla u \right) - \lambda(u - u_0) \quad \text{in } \Omega \times \mathbb{R}_+, \quad (45)$$

with homogeneous Neumann boundary conditions.

7. EXPERIMENTAL RESULTS

This section presents simulation results where Huber, entropic, total variation, and improved entropic gradient descent flows are applied to enhance images corrupted by Gaussian and Laplacian noise.

The performance of a filter clearly depends on the filter type, the properties of signals/images, and the characteristics of the noise. The choice of criteria by which to measure

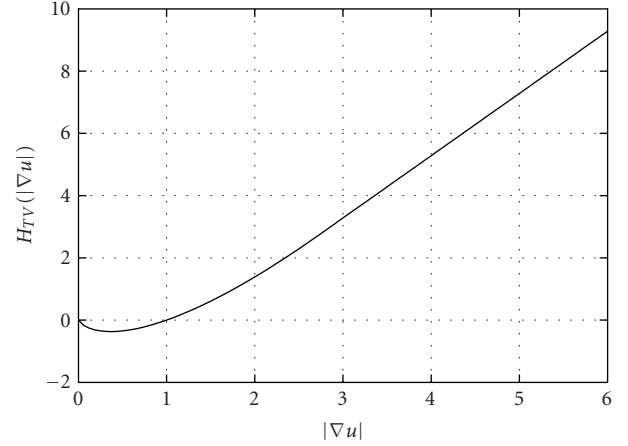


FIGURE 8: Improved entropic Lagrangian.

the performance of a filter presents certain difficulties, and only gives a partial picture of reality. To assess the performance of the proposed denoising methods, a mean square error (MSE) between the filtered and the original image is evaluated and used as a quantitative measure of performance of the proposed techniques. The regularization parameter (or Lagrange multiplier) λ for the proposed gradient descent flows is chosen to be proportional to signal-to-noise ratio (SNR) in all the experiments.

In order to evaluate the performance of the proposed gradient descent flows in the presence of Gaussian noise, the image shown in Figure 9a has been corrupted by Gaussian white noise with SNR = 4.79 dB. Figure 9 displays the results of filtering the noisy image shown in Figure 9b by Huber with optimal $k = 1.345$, entropic, total variation, and improved entropic gradient descent flows. Qualitatively, we observe that the proposed techniques are able to suppress Gaussian noise while preserving important features in the image. The resulting MSE computations are depicted in Table 1. Figure 10 depicts another numerical example, and it clearly illustrates that the improved entropic flow performs the best. This fact is consistent with a variety of images used for experimentation.

The Laplacian noise is somewhat heavier than the Gaussian noise. Moreover, the Laplace distribution is similar to Huber's least favorable distribution [26] at least in the tails. To demonstrate the application of the proposed gradient descent flows to image denoising, qualitative and quantitative comparisons are performed to show a much improved performance of these techniques. Figure 11b shows a noisy image contaminated by Laplacian white noise with SNR = 3.91 dB. The MSE's results obtained by applying the proposed techniques to the noisy image are shown in Table 2. Note that from Figure 11, it is clear that the improved entropic gradient descent flow outperforms the other flows in removing Laplacian noise. Comparison of these images clearly indicates that the improved entropic gradient descent flow preserves well the image structures while removing heavy-tailed noise.



FIGURE 9: Filtering results for Gaussian noise (Lena image): (a) original image, (b) noisy image, (c) Huber, (d) entropic, (e) total variation, and (f) improved entropic.

TABLE 1: MSE's computations for Gaussian noise.

PDE	MSE		
	SNR = 4.79	SNR = 3.52	SNR = 2.34
Huber	234.1499	233.7337	230.0263
Entropic	205.0146	207.1040	205.3454
Total variation	247.4875	263.0437	402.0660
Improved entropic	121.2550	137.9356	166.4490

The relative error versus the iteration number is illustrated in Figure 12, where the convergence of the improved entropic flow is clearly demonstrated. The stopping criterion for the proposed flow is $\|u^{(n+1)} - u^{(n)}\|_2 / \|u^{(n)}\|_2 < \epsilon$, where ϵ is sufficiently small and n denotes the n th iteration.

8. DISCUSSIONS AND CONCLUSIONS

In this paper, we have explored a connection between maximum a posteriori estimation and the variational formulation based on the minimization of a given variational integral subject to some noise constraints. Huber gradient descent flow minimizes the Huber variational integral subject to some noise constraints. This filter behaves as the total variation anisotropic diffusion for small gradient magnitudes and as the isotropic diffusion for large gradient magnitudes. On the other hand, an improved entropic gradient descent flow which is derived from the maximum entropy principle was proposed. It minimizes a hybrid functional between the negentropy variational integral and Huber integral, subject to some noise constraints. The proposed gradient descent flows have been applied to enhance images corrupted

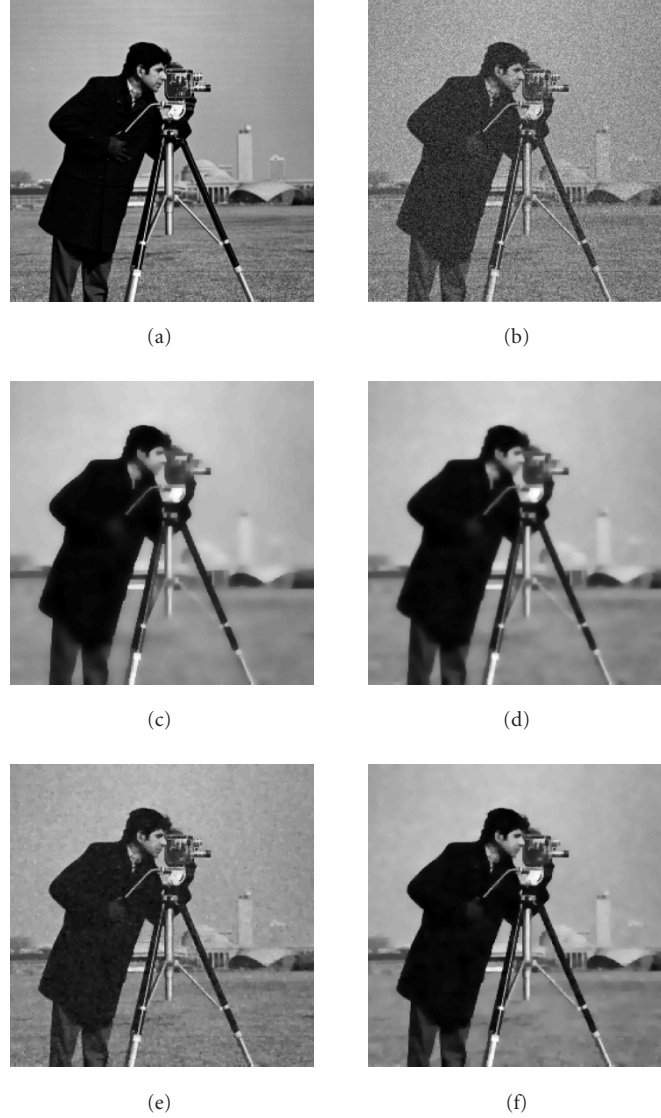


FIGURE 10: Filtering results for Gaussian noise (Cameraman image): (a) original image, (b) noisy image, (c) Huber, (d) entropic, (e) total variation, and (f) improved entropic.

by Gaussian as well as Laplacian noise, and it has been shown that these proposed techniques preserve details well while removing noise.

APPENDICES

A. DERIVATION OF GRADIENT FLOWS

The first variation of the functional $\mathcal{F}(u) = \int_{\Omega} F(|\nabla u|) dx$ in the direction of v is given by

$$\begin{aligned} \delta \mathcal{F}(u; v) &= \left. \frac{d}{d\epsilon} \mathcal{F}(u + \epsilon v) \right|_{\epsilon=0} \\ &= \int_{\Omega} \left(\frac{F'(|\nabla u|)}{|\nabla u|} \nabla u \cdot \nabla v \right) dx. \end{aligned} \quad (\text{A.1})$$

The identity

$$\operatorname{div}(v \nabla u) = \operatorname{div}(\nabla u) v + \nabla u \cdot \nabla v \quad (\text{A.2})$$

yields

$$\begin{aligned} &\int_{\Omega} \frac{F'(|\nabla u|)}{|\nabla u|} \nabla u \cdot \nabla v dx \\ &= - \int_{\Omega} \operatorname{div} \left(\frac{F'(|\nabla u|)}{|\nabla u|} \nabla u \right) v dx \\ &\quad + \int_{\Omega} \operatorname{div} \left(v \frac{F'(|\nabla u|)}{|\nabla u|} \nabla u \right) dx. \end{aligned} \quad (\text{A.3})$$

Using the divergence theorem for a vector field \mathbf{w} ,

$$\int_{\Omega} \operatorname{div}(\mathbf{w}) dx = \int_{\partial\Omega} \mathbf{w} \cdot \boldsymbol{\nu} ds, \quad (\text{A.4})$$

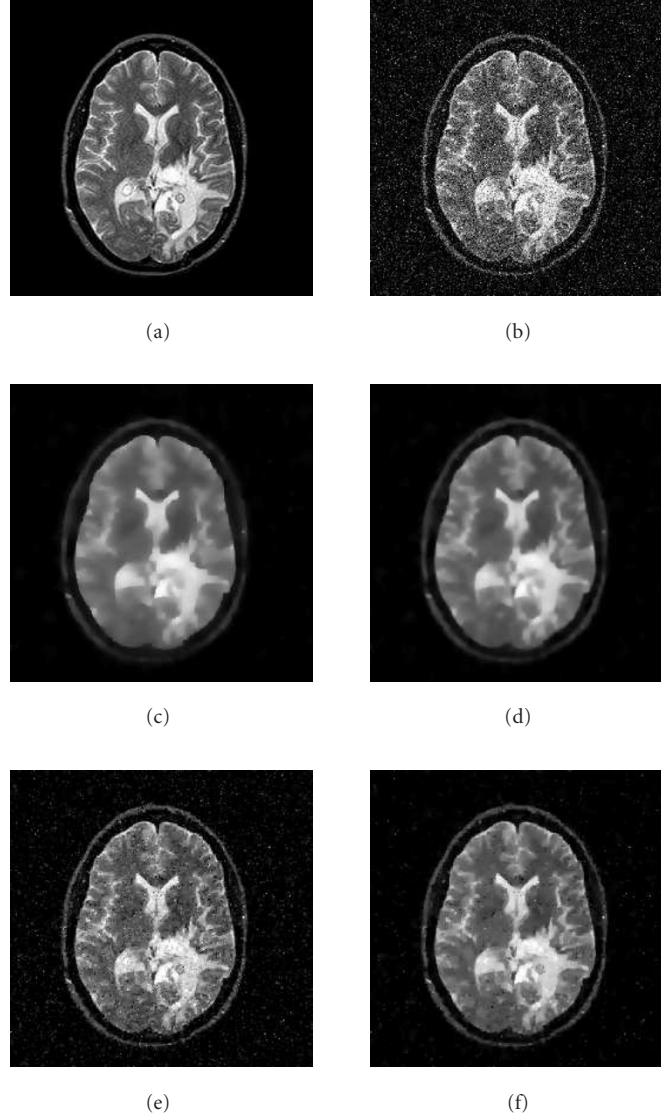


FIGURE 11: Filtering results for Laplacian noise: (a) original image, (b) noisy image, (c) Huber, (d) entropic, (e) total variation, and (f) improved entropic.

TABLE 2: MSE computations for Laplacian noise.

PDE	MSE		
	SNR = 6.33	SNR = 3.91	SNR = 3.05
Huber	237.7012	244.4348	248.4833
Entropic	200.5266	211.4027	217.3592
Total variation	138.4717	176.1719	213.1221
Improved entropic	104.4591	170.2140	208.8639

where $\boldsymbol{\nu}$ is the outward unit normal vector (field) on $\partial\Omega$ (the boundary of Ω) and ds is an area element. Therefore,

$$\int_{\Omega} \operatorname{div} \left(\nu \frac{F'(|\nabla u|)}{|\nabla u|} \nabla u \right) d\mathbf{x} = \int_{\partial\Omega} \nu \frac{F'(|\nabla u|)}{|\nabla u|} \nabla u \cdot \boldsymbol{\nu} ds. \quad (\text{A.5})$$

If we assume homogenous Neumann boundary conditions

$$\nabla u \cdot \boldsymbol{\nu} = u_{\boldsymbol{\nu}} = \frac{\partial u}{\partial \boldsymbol{\nu}} = 0, \quad (\text{A.6})$$

then the first variation of \mathcal{F} is reduced to

$$\delta \mathcal{F}(u; v) = - \int_{\Omega} \operatorname{div} \left(\frac{F'(|\nabla u|)}{|\nabla u|} \nabla u \right) v d\mathbf{x} \quad \forall v \in X \quad (\text{A.7})$$

which concludes the proof.

B. NUMERICAL IMPLEMENTATION OF GRADIENT FLOWS

A numerical implementation of the PDE given by (19) is performed using an explicit scheme in time and location as

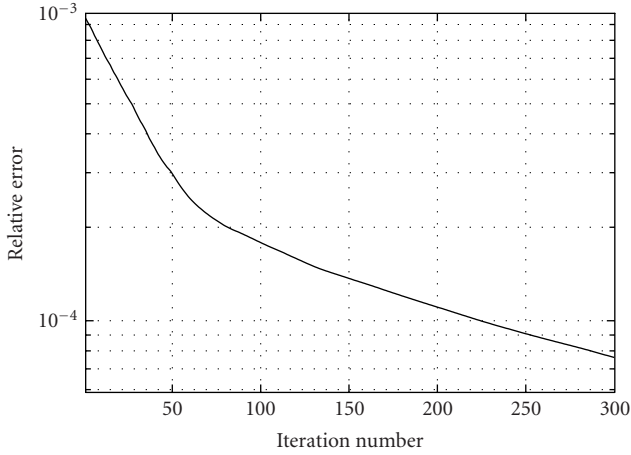


FIGURE 12: Improved entropic flow: relative error versus iteration number.

follows. Let $u_{i,j}^n$ be the approximation of $u(x, y, t)$ on a grid $(i\Delta x, j\Delta y, n\Delta t)$. For simplicity, we assume that $\Delta x = \Delta y = h$. Denote by

$$D_{\pm}^x u = \pm \frac{u_{i\pm 1,j}^n - u_{i,j}^n}{h}, \quad D_{\pm}^y u = \pm \frac{u_{i,j\pm 1}^n - u_{i,j}^n}{h} \quad (\text{B.1})$$

the matrices of column differences and row differences, respectively (i.e., backward and forward differences).

Similarly, the central differences are given by

$$D_c^x u = \frac{u_{i+1,j}^n - u_{i-1,j}^n}{2h}, \quad D_c^y u = \frac{u_{i,j+1}^n - u_{i,j-1}^n}{2h}. \quad (\text{B.2})$$

The operator $\text{div}(g(|\nabla u|)\nabla u)$ on the right-hand side of (19) is discretized using an upwind scheme as follows:

$$\begin{aligned} & \text{div}(g(|\nabla u|)\nabla u) \\ &= \frac{1}{h} \left\{ D_-^x \left(g \left[\left(\left(\frac{D_+^x u}{h} \right)^2 + \left(\min \text{mod} \left(\frac{D_+^y u}{h}, \frac{D_-^y u}{h} \right) \right)^2 \right)^{1/2} \right] \left(\frac{D_+^x u}{h} \right) \right) \right. \\ & \quad \left. + \frac{1}{h} \left\{ D_-^y \left(g \left[\left(\left(\frac{D_+^y u}{h} \right)^2 + \left(\min \text{mod} \left(\frac{D_+^x u}{h}, \frac{D_-^x u}{h} \right) \right)^2 \right)^{1/2} \right] \left(\frac{D_+^y u}{h} \right) \right) \right. \right. \\ & \quad \left. \left. + \left(\min \text{mod} \left(\frac{D_+^x u}{h}, \frac{D_-^x u}{h} \right) \right)^2 \right)^{1/2} \right] \left(\frac{D_+^y u}{h} \right) \right\} \right\}, \quad (\text{B.3}) \end{aligned}$$

where $\min \text{mod}$ is a function that returns the argument with the smallest absolute value when all the arguments are of the same sign and zero otherwise. The $\min \text{mod}$ function is a limiter whose goal is to prevent oscillations while maintaining

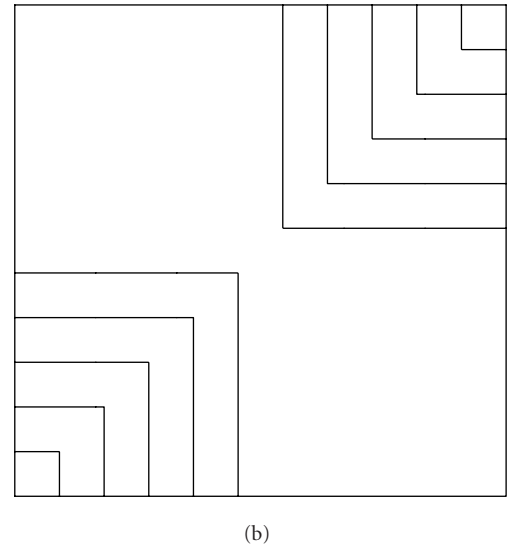
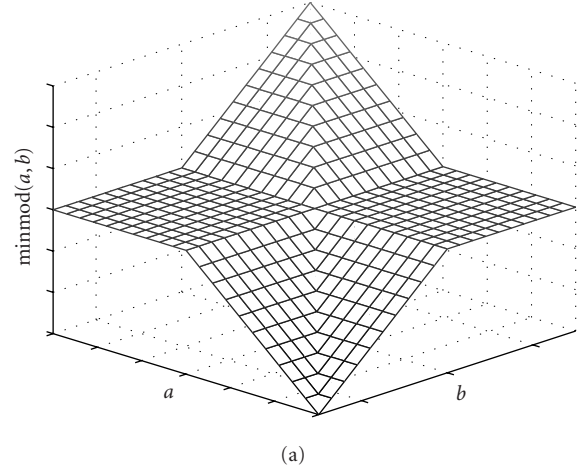


FIGURE 13: Illustration of $\min \text{mod}$ function: (a) 3D plot of the $\min \text{mod}$ function, (b) level curves.

the order of accuracy of the method, and it is defined as

$$\begin{aligned} \min \text{mod}(a, b) &= \left(\frac{\text{sign}(a) + \text{sign}(b)}{2} \right) \min(|a|, |b|) \\ &= \min(\max(a, 0), \max(b, 0)) \\ & \quad + \max(\min(a, 0), \min(b, 0)) \quad (\text{B.4}) \\ &= \begin{cases} 0 & \text{if } ab \leq 0, \\ a & \text{if } |a| \leq |b|, ab > 0, \\ b & \text{if } |a| > |b|, ab > 0. \end{cases} \end{aligned}$$

Figure 13a depicts the gradient vector field of the $\min \text{mod}$ function (the background color indicates the value of the $\min \text{mod}$ function), and Figure 13b illustrates its contours.

ACKNOWLEDGMENTS

The authors would like to thank the anonymous reviewers for helpful and constructive comments. This work was in part supported by US Air Force Office of Scientific Research Grant no. F49620-98-1-0190, and by NATO Collaborative Linkage Grant no. 980107. Part of this work was conducted when H. Krim was visiting INRIA as an invited Professor.

REFERENCES

- [1] D. Mumford and J. Shah, "Optimal approximations by piecewise smooth functions and associated variational problems," *Comm. Pure Appl. Math.*, vol. 42, no. 5, pp. 577–685, 1989.
- [2] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 12, no. 7, pp. 629–639, 1990.
- [3] L. Alvarez, P.-L. Lions, and J.-M. Morel, "Image selective smoothing and edge detection by nonlinear diffusion. II," *SIAM J. Numer. Anal.*, vol. 29, no. 3, pp. 845–866, 1992.
- [4] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, no. 1–4, pp. 259–268, 1992.
- [5] L. Alvarez and L. Mazorra, "Signal and image restoration using shock filters and anisotropic diffusion," *SIAM J. Numer. Anal.*, vol. 31, no. 2, pp. 590–605, 1994.
- [6] J. Weickert, *Anisotropic Diffusion in Image Processing*, Teubner Verlag, Stuttgart, Germany, 1998.
- [7] J.-M. Morel and S. Solimini, *Variational Methods in Image Segmentation*, Birkhäuser, Boston, Mass, USA, 1995.
- [8] I. Pollak, A. S. Willsky, and H. Krim, "Image segmentation and edge enhancement with stabilized inverse diffusion equations," *IEEE Trans. Image Processing*, vol. 9, no. 2, pp. 256–266, 2000.
- [9] R. Deriche and O. Faugeras, "Les EDP en traitement des images et vision par ordinateur," *Traitement du Signal*, vol. 13, no. 6, pp. 59, 1996.
- [10] P. Kornprobst, R. Deriche, and G. Aubert, "Image sequence analysis via partial differential equations," *J. Math. Imaging Vision*, vol. 11, no. 1, pp. 5–26, 1999.
- [11] C. Samson, L. Blanc-Feraud, G. Aubert, and J. Zerubia, "A variational model for image classification and restoration," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, no. 5, pp. 460–472, 2000.
- [12] G. Aubert and L. Vese, "A variational method in image recovery," *SIAM J. Numer. Anal.*, vol. 34, no. 5, pp. 1948–1979, 1997.
- [13] H. Krim and I. C. Schick, "Minimax description length for signal denoising and optimized representation," *IEEE Transactions on Information Theory*, vol. 45, no. 3, pp. 898–908, 1999.
- [14] A. Ben Hamza and H. Krim, "Image denoising: a nonlinear robust statistical approach," *IEEE Trans. Signal Processing*, vol. 49, no. 12, pp. 3045–3054, 2001.
- [15] A. Yezzi, "Modified curvature motion for image smoothing and enhancement," *IEEE Trans. Image Processing*, vol. 7, no. 3, pp. 345–352, 1998.
- [16] P. Charbonnier, L. Blanc-Feraud, G. Aubert, and M. Barlaud, "Deterministic edge-preserving regularization in computed imaging," *IEEE Trans. Image Processing*, vol. 6, no. 2, pp. 298–311, 1997.
- [17] M. Cetin and W. C. Karl, "Feature-enhanced synthetic aperture radar image formation based on nonquadratic regularization," *IEEE Trans. Image Processing*, vol. 10, no. 4, pp. 623–631, 2001.
- [18] A. Ben Hamza and H. Krim, "A variational approach to maximum a posteriori estimation for image denoising," in *Energy Minimization Methods in Computer Vision and Pattern Recognition*, vol. 2134 of Lecture Notes in Computer Science, pp. 19–34, Springer-Verlag, New York, NY, USA, 2001.
- [19] T. F. Chan and C.-K. Wong, "Total variation blind deconvolution," *IEEE Trans. Image Processing*, vol. 7, no. 3, pp. 370–375, 1998.
- [20] T. F. Chan, S. Osher, and J. Shen, "The digital TV filter and nonlinear denoising," *IEEE Trans. Image Processing*, vol. 10, no. 2, pp. 231–241, 2001.
- [21] M. Giaquinta and S. Hildebrandt, *Calculus of Variations I: The Lagrangian Formalism*, Springer-Verlag, Berlin, Germany, 1996.
- [22] T. F. Chan and P. Mulet, "On the convergence of the lagged diffusivity fixed point method in total variation image restoration," *SIAM J. Numer. Anal.*, vol. 36, no. 2, pp. 354–367, 1999.
- [23] K. Ito and K. Kunisch, "Restoration of edge-flat-grey scale images," *Inverse Problems*, vol. 16, no. 4, pp. 909–928, 2000.
- [24] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 6, no. 6, pp. 721–741, 1984.
- [25] Y.-L. You, W. Xu, A. Tannenbaum, and M. Kaveh, "Behavioral analysis of anisotropic diffusion in image processing," *IEEE Trans. Image Processing*, vol. 5, no. 11, pp. 1539–1553, 1996.
- [26] P. Huber, *Robust Statistics*, John Wiley & Sons, New York, NY, USA, 1981.
- [27] M. J. Black, G. Sapiro, D. H. Marimont, and D. Heeger, "Robust anisotropic diffusion," *IEEE Trans. Image Processing*, vol. 7, no. 3, pp. 421–432, 1998.
- [28] A. Ben Hamza and H. Krim, "Robust influence functionals for image filtering," in *Proc. IEEE International Conference on Image Processing (ICIP '03)*, vol. 3, pp. 361–364, Barcelona, Spain, September 2003.
- [29] G. Gallager, *Information Theory and Reliable Communication*, John Wiley & Sons, New York, NY, USA, 1968.
- [30] H. Krim, "On the distributions of optimized multiscale representations," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP '97)*, vol. 5, pp. 3673–3676, Munich, Germany, April 1997.

A. Ben Hamza received his Ph.D. degree in electrical engineering from North Carolina State University in 2003, where he worked on variational methods in imaging and computer vision, and information theory. Prior to joining Concordia Institute for Information Systems Engineering, Concordia University, he was a Postdoctoral Associate at Duke University, affiliated with both the Department of Electrical and Computer Engineering and the Fitzpatrick Center for Photonics and Communications Systems. His research interests include geometric and topological modeling for 3D graphics, wavelets, computational imaging and vision, information-theoretic measures, and human tracking using pyroelectric infrared sensors.



Hamid Krim received his B.S., M.S., and Ph.D. degrees in electrical engineering from the University of Southern California, the University of Washington, and Northeastern University, respectively. As a member of the technical staff at AT&T Bell Labs, he worked in the area of telephony and digital communication systems/subsystems. In 1991, he became a US National Science Foundation Postdoctoral Scholar at the Foreign Centers of Excellence (LSS Supelec/University of Orsay, Paris, France). He subsequently joined the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, Massachusetts, as a Research Scientist performing/supervising research in his area of interest, and later became a faculty member in the Electrical and Computer Engineering Department, North Carolina State University in Raleigh in 1998. He is an original contributor and now an Affiliate of the Center for Imaging Science sponsored by the US Army. He is also a recipient of the US National Science Foundation Career Young Investigator Award. He is a Senior Member of the IEEE and a Member of the IEEE Computer Society.



Josiane Zerubia has been a Permanent Research Scientist at INRIA since 1989. She has been the Director of Research since July 1995. Since January 1998, she has been in charge of a research group working on remote sensing (ARIANA). She has been an Adjunct Professor at Sup'Aero (ENSAE) in Toulouse since 1999. Before, she was at the University of Southern California in Los Angeles as a Postdoc. She also worked as a Researcher for the University of Nice and CNRS from 1984 to 1988, and in the Research Lab of Hewlett Packard from 1982 to 1984. She got the M.S. degree from ENSIEG, Grenoble, France, in 1981, and a Doctor Engineer degree in 1986, a Ph.D. degree in 1988, and a "Habilitation" degree in 1994, all from UNSA, France. She is a Fellow of the IEEE. She was part of the IEEE IMDSP Technical Committee (SP Society) from 1997 to 2003 and Associate Editor of IEEE Transactions on IP from 1998 to 2002. She has been Member-at-Large of the Board of Governors of IEEE SP Society since 2002, Area Editor of IEEE Transactions on IP since 2003, and Guest Coeditor of a special issue of IEEE Transactions on PAMI in 2003. She has also been a Member of the Editorial Board of the French Society for Photogrammetry and Remote Sensing (SFPT) since 1998. Her current research interest is in image processing using probabilistic models or variational methods. She also works on parameter estimation and optimization techniques.

